# Data-driven approach for isolated PV shading fault diagnosis based on experimental I-V curves analysis

Siwar Fadhel, Anne Migan-Dubois, Claude Delpha, Demba Diallo, Imen Bahri, Mohamed Trabelsi, Mohamed Faouzi Mimouni

## HAL Id: hal-01702863
## https://hal.science/hal-01702863

Submitted on 12 Mar 2020

# Data-driven Approach for Isolated PV Shading Fault Diagnosis Based on Experimental I-V curves Analysis

S. Fadhel[1,2,3], A. Migan[1], C. Delpha[4], D. Diallo[1], I. Bahri[1] , M. Trabelsi[3], M. Faouzi-Mimouni[3]

1: Group of Electrical Engineering of Paris
UMR 8507 CNRS, CentraleSupelec, UPMC, Univ. Paris Sud, Univ. Paris Saclay
Gif Sur Yvette, France
2: National engineering school of Sousse, Univ. of Sousse, Tunisia
3: Research unit of Study of Industrial Systems and Renewable Energy,
National engineering school of Monastir, Univ. of Monastir, Tunisia
4: Laboratory of Signal and Systems
UMR 8506 CNRS, CentraleSupelec, Univ. Paris Sud
Gif Sur Yvette, France

*Abstract*—**This paper deals with a data-driven fault diagnosis method for photovoltaic (PV) system. The proposed method is based on the Principal Component Analysis (PCA) to detect and identify different shading types. The PCA uses the current- voltage (I-V) curves that are experimentally determined for a monocrystalline PV module of 250Wc. The experimental tests were carried out for several shading patterns covering the PV cells. For the diagnosis process, three features (current, voltage and power of PV module) are extracted for each test to build the database which is then analyzed through the PCA algorithm. Simulation results using the experimental data, prove the efficiency of the proposed method in terms of discrimination. The healthy data are clearly separated from the faulty ones despite sudden irradiation variations.**

*Keywords— PV cells, Shadow faults ; I-V curves; Principal Component Analysis; Fault diagnosis*

## I. INTRODUCTION

Solar energy has received much attention in recent years, as an alternative form of electricity source using the photovoltaic (PV) technology in both residential and commercial applications, reaching a global installed capacity of 296 GW in the world, at the end of 2016 [1]. In the literature, PV modules present an overall performance degradation quantified by an annual average degradation rate of 0.8% per year [2]. The environment is one of the contributing factors that directly affect the PV module performances. In this context, a great number of recent researches have been focused on environment factors effects on PV performances such as soiling and shading that cover the PV cells [3]. According to [4], partial shading can lead to more than 10-20% of annual reduction in power production in residential installations. Apart from power losses, the presence of localized shading on PV modules leads to an overheating of the shaded PV cells despite the presence of bypass diodes. The IR thermography such in [5,6] proves the presence of hot spot zones of the shaded PV cells. The temperature increase in these zones reduces considerably the PV module lifetime and can damage the shaded cells [7]. Fault diagnosis in such systems is thus necessary to improve their productivity, efficiency and safety. Several techniques have been developed to detect faults in PV systems. These techniques can be classified into two major categories: model-based approaches and process history-based approaches also named data-driven approaches. Model-based approaches generally use an analytical model of the PV system. This model is used to estimate the parameters, which will be compared to the measured ones obtained from real time data to generate residues for diagnosis purposes. If the model can capture the main physics of the system, these methods are efficient for shading detection [8-9]. Process history-based approaches use available data of the process and rely often on computational intelligence and machine learning methods. Among this category of techniques, those based on signal analysis such as time-domain reflectometry (TDR) [10,11] are used to detect and identify open-circuit fault. Spread spectrum time-domain reflectometry (SSTDR) techniques are used to detect catastrophic faults, ground-faults and PV arc faults [12,13]. However these are not used to detect and identify shading faults. Another group of data-driven approaches is based on the analysis of the I-V characteristic for PV module, string or array. The study in [14] detects the activation of bypass diodes indicating the presence of shading, by calculating the first and the second derivatives of (I-V). In [6], the detection of the shading is done through the comparison of I-V curves in normal and shaded operations. The methods presented in [6,14] are simple and can detect shading faults, but they are not able to identify and classify the type of shading patterns. The shading faults can also be detected with artificial neural network (ANN) and fuzzy classifier system as described in [15,16]. The main disadvantage of these techniques is that the rules must be designed for each system.

In this paper, authors propose a shading fault detection method based on the analysis of the features extracted from the I-V curve using Principal component analysis (PCA).

Principal component analysis (PCA) is a data-driven tool that has been proved in several studies to be effective and powerful for the diagnosis in different applications [17-19]. To

our knowledge, PCA has not yet been used for PV systems diagnosis.

This paper is outlined as follows. Section II presents a brief description of the experimental setup used in this study. Section III introduces the PCA. The test protocol is detailed and the results are presented and discussed. The final section concludes the paper.

## II. EXPERIMENTAL SET-UP DESCRIPTION

The experimental setup is installed at the SIRTA [20] meteorological station in Palaiseau, France. A picture and a diagram describing the instrumentation are displayed in Fig.1. The experimental photovoltaic generator used in this study, is a monocrystalline PV module, whose specifications are listed in Table I under standard test condition (STC: solar irradiation of 1000 W/m$^2$, module temperature of 25°C). The PV module under test is composed of 60 PV cells connected in series and gathered into three sub-strings of 20 PV cells. This module is equipped with three bypass diodes; each one is mounted in anti-parallel to protect a PV sub-string. The structure of Fig.1 comprises:

- the PV module under test (FL 6-250MBP by FranceWatts),

- a reference cell (RG100 by SOLEMS) used to measure the solar irradiance captured by the PV module area,

- a 4-wire Pt100 probe used to measure the temperature on the back face of the PV module,

- a variable load (Programmable DC electronic load chroma 63600) used to sweep the I-V characteristic,

- a data acquisition system and a computer for supervision and data visualization using LabVIEW.

TABLE I.     FL60-250MBP MODULE SPECIFICATIONS AT STC

| Symbol | Quantity | Value |
|--------|----------|-------|
| $P_{mpp}$ | Maximum Power | 250 Wc |
| $I_{mpp}$ | Current at $P_{mpp}$ | 8.21 A |
| $V_{mpp}$ | Voltage at $P_{mpp}$ | 30.52 V |
| $I_{sc}$ | Short-circuit Current | 8.64 A |
| $V_{oc}$ | Open-circuit voltage | 37.67 V |

## III. FAULT DETECTION BASED ON PCA METHOD

### A. PCA Principle

Principal Component Analysis (PCA) is a multivariate statistical method, which seeks in the multidimensional space of process variables, the most interesting dimensions to re-express the multivariate dataset built from a large number of measurements collected at different sampling times. The new dimensions are uncorrelated so the reduced subspace behaves like a denoising filter and keeps the underlying "latent structures" in the data.

This subspace is called the principal subspace or the 'representation' subspace. Its complementary subspace into which noises and outliers are rejected is termed the residual subspace. Analytically, PCA seeks orthogonal directions, which maximise the variance of projections of the dataset points. These directions are called principal components (PCs), each one being characterised by a loading vector and a score vector. Each PC is aligned in a direction corresponding to the largest variance of the data, starting with the first PC. Principal components are therefore ordered from the most energised (with highest eigenvalue) to the less energised (with the lowest eigenvalue). Based on stop criteria the most energised PCs span the principal subspace while the remaining ones span the residual subspace.

The solution to the PCA problem is given by the eigenvector decomposition of the data covariance matrix [21]. The obtained eigenvectors are the loading vectors, and the scores are projections of data points onto these vectors. From the data collected in the normal (healthy) operating mode of the process, PCA is applied and the loading vectors are used as references to design the process model. The scores or their statistical distributions can be also used to design the implicit model. When new data are collected, they are projected in the subspaces (principal or/and residual). The deviation from the reference is measured and analysed to assess the fault occurrence.

The data consist of measurements at *N* different sampling times of *m* physical variables termed the features. The time points represent the observations. PCA uses a linear combination of the original variables to build the new variables while keeping maximum variance information. The first principal components, which span the principal subspace, are given by the first *l* dominant eigenvectors of the data covariance matrix. They are associated to the *l* highest eigenvalues. The last non-retained eigenvectors (*m − l*) define the residual subspace. In the principal subspace containing the most significant variations, the eigenvectors are denoted loading vectors and the new data which are projections of the system data on the loading vectors are called principal component scores.

Let us consider *N* observations of *m* variables arranged into the original data matrix $X_{[N \times m]}$ given by:

$$X_{[N \times m]} = [x_1, ..., x_k, .., x_N] \qquad (1)$$

where $x_{j(j=1,.N)}$ is the $j^{th}$ variable.

The first consists of centering (zero mean) and reducing each variable such as in (2) for each observation of $x_{j(j=1,.N)}$:

$$(x_j)_c(k) = \frac{x_j(k) - (\overline{x}_j)_m}{(\sigma_j)_m} \qquad (2)$$

where $(x_j)_c$ is the centered and reduced variable, $(\bar{x}_j)_m$ and $(\sigma_j)_m$ are respectively the mean value and the standard deviation of $x_{j(j=1,..N)}$.

We can therefore define the new data matrix as

$$(X_c)_{[N \times m]} = [(x_1)_c,...,(x_k)_c,...,(x_N)_c]$$

The covariance matrix is then calculated as:

$$C = \frac{1}{N-1} X_c^T X_c \qquad (3)$$

Let us denote $P$ the column matrix of loading vectors, which are arranged in the descendent order of their corresponding eigenvalues. The principal component scores are finally obtained by the projection of the original data centered and reduced on the new basis P. The matrix $T_{[N \times m]}$ of the principal component scores is defined by:

$$T_{[N \times m]} = X_{[N \times m]} P_{[m \times m]} = [t_1,...,t_k,..,t_N] \qquad (4)$$
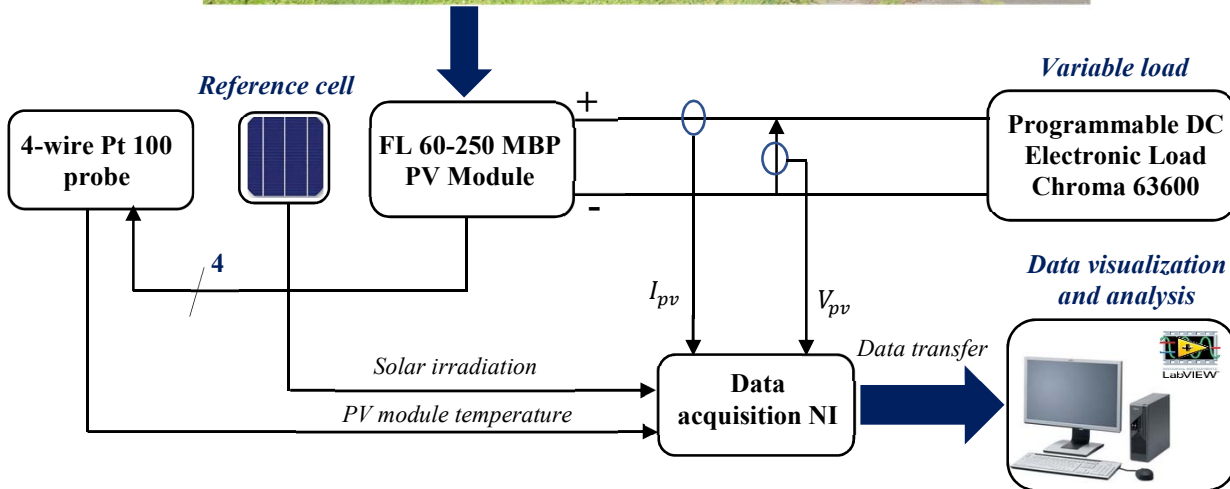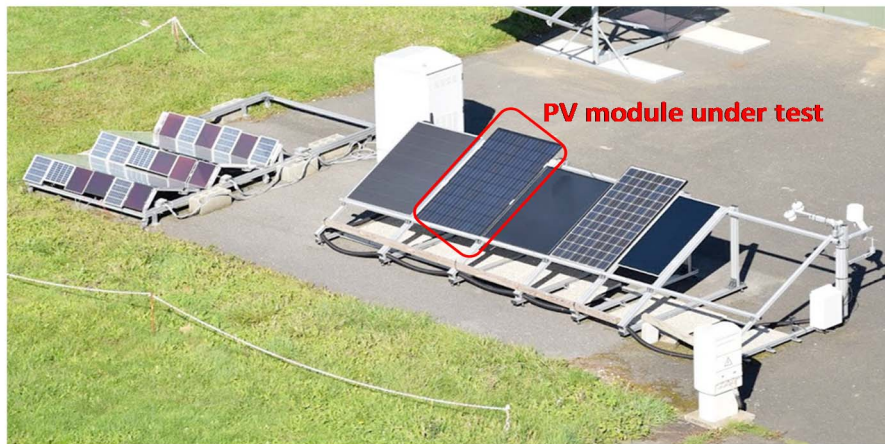


Fig. 1: Layout of the experimental setup

### B. Test protocol for PCA data processing

The experimental tests have been performed under several operating conditions:

- healthy mode with and without soiling that is considered as an external source of perturbation for the healthy mode,

- and faulty modes with different shading patterns applied on the PV cells as shown in Fig.3 and Fig.5.

A total of 8 sets of experimental tests have been conducted. The corresponding I-V curves are illustrated in Fig. 2, 4 and 6.

The mean duration between two measurements of one I-V characteristic is less than one minute. Despite this short duration, the irradiation can change significantly. Each I-V curve consists

of 101 samples. The behavior of all experimental characteristics differs according to the shaded pattern (row level, column level, number of shaded cells) and to the climatic variations (temperature and solar irradiation).

For each I-V curve (Fig.2, 4 and 6), the irradiation displayed in the legend is the one measured with the reference cell.

### 1) Scenario 1: Healthy test condition

Fig.2 illustrates the experimental I-V curves when the PV module operates in normal condition considered as reference. Initially, the PV panel was dusty and some cells were covered with bird soiling. To show the effect of the soiling, the PV module area was manually cleaned with water. After cleaning the I-V curve was also measured while the PV panel was still wet. Finally after drying the PV panel, the I-V curve has also been measured. The I-V curves show the effect of the presence of soiling which results in a little decrease of the PV current as it reduces the quantity of the solar irradiation received by PV cells. The amount of this decrease is proportional to the quantity of soiling accumulation on PV cells. However, the I-V curves have the same shape before and after cleaning.
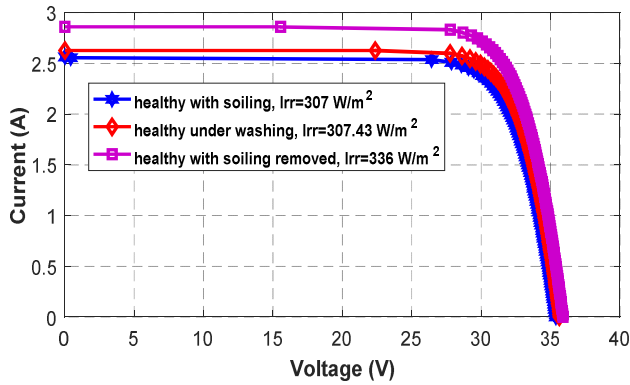


Fig. 2: I-V curves under healthy (normal) conditions

### 2) Scenario 2: shading patterns 1, 2 and 3

A survival blanket is used to cover the cells during the tests. The first three shading patterns are illustrated in Fig.3:

- ✓ Shading pattern 1: four rows of the module are shaded,

- ✓ Shading pattern 2: six PV cells of the first column are shaded,

- ✓ Shading pattern 3: six PV cells of the first column and six additional PV cells of the second column are shaded

In general, under the shading conditions, the current generated by the unshaded cells flows through the bypass diodes instead of the shaded cells. The activation of these diodes prevents the shaded cells from heating up and becoming hot spots. The deviations in the I-V curves of Fig.4 clearly reveal the activation of one or two bypass diodes.
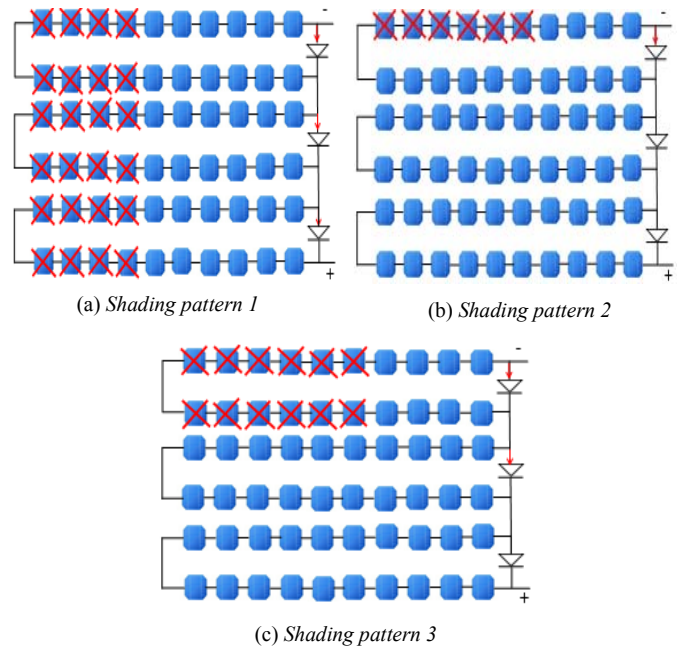


(a) *Shading pattern 1*  (b) *Shading pattern 2*



(c) *Shading pattern 3*

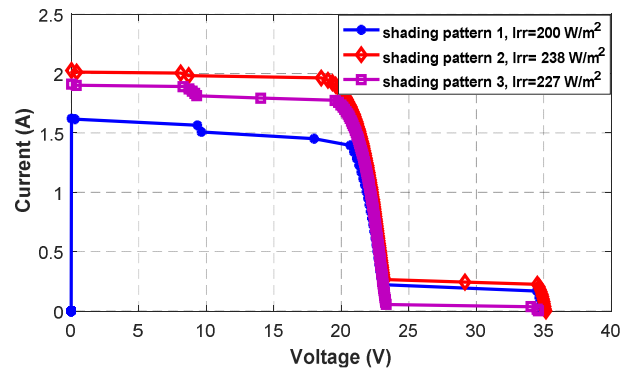Fig. 3: Shading patterns for scenario 2



Fig. 4: I-V curves for scenario 2

### 3) Scenario 3: shading patterns 4 and 5

Fig.5 illustrates the case of shaded patterns 4 and 5. They correspond respectively to 18 and 24 shaded cells. The impact is clearly illustrated by the deviations in the I-V curves as shown in Fig.6.
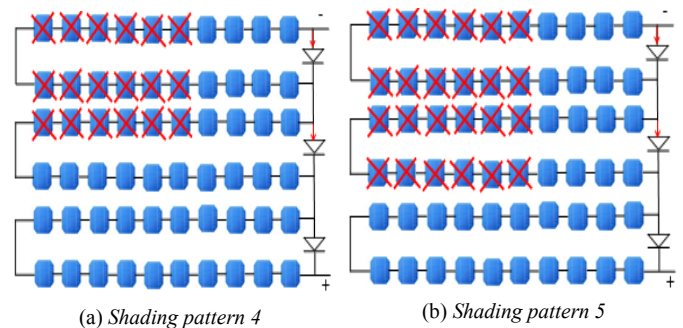


(a) *Shading pattern 4*  (b) *Shading pattern 5*
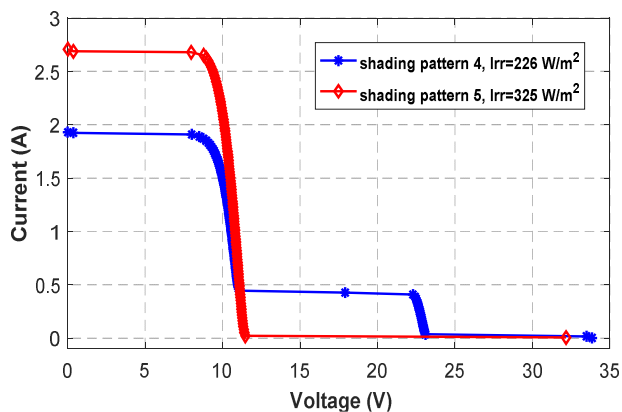
Fig. 5: Shading patterns for scenario 3

Fig.6: I-V curves for scenario 3

## C. Results and discussions

To perform the PCA algorithm, we first build our database matrix. $X_{[808 \times 3]} = \begin{bmatrix} I & V & P \end{bmatrix}$ is composed of 808 observations. 101 samples for each of the 8 experiments with 3 variables (the current, the voltage and the power of the PV module) are enough to sweep each I-V curve of the module under study. The number of the samples required to extract the characteristics is selected according to the size of the PV system and to the mismatching conditions in order to clearly show the deviations and the inflection points on the I-V curve. The data is collected and processed under Matlab®.

To select the scores that include the maximum data variance, we plot in Fig.7 the relative contribution of each eigenvalue. As it can be deduced, the first and the second scores represent 97% of the cumulated variance.
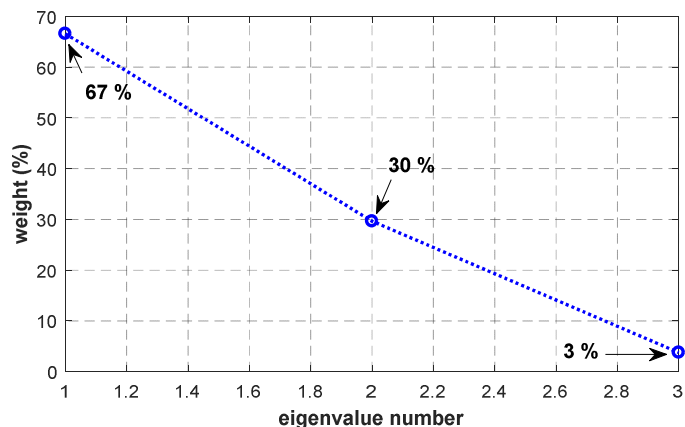


Fig.7: Contribution (in percentage) of the eigenvalues

The data projection into the 2-dimensionnal subspace spanned with the first and the second principal components (PC1 and PC2) gives the data scatter illustrated in Fig.8. Another graphical representation of the data in PCA subspace is proposed in Fig.9.

Despite the variation of the solar irradiation during the measurement, we can separate the data into three different classes: C1 that gathers healthy data including slight soiling, C2 that represents the first shading patterns 1, 2 and 3 and C3 that

corresponds to the second shading patterns 4 and 5. This classification is consistent with the I-V curves obtained for the tests. Moreover, one can notice that the three classes are clearly separated.
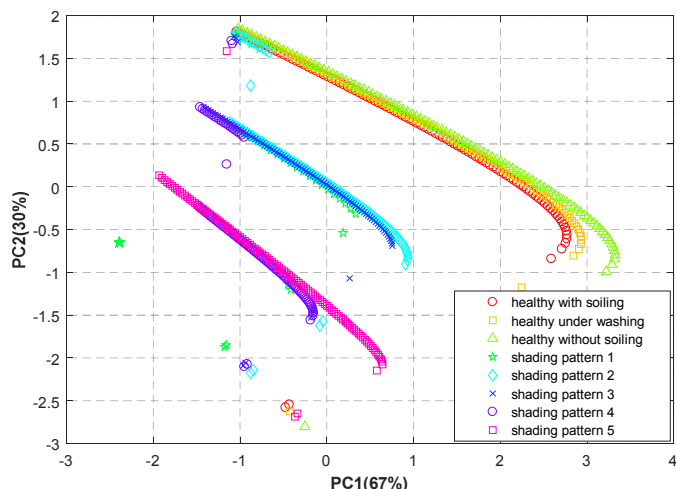


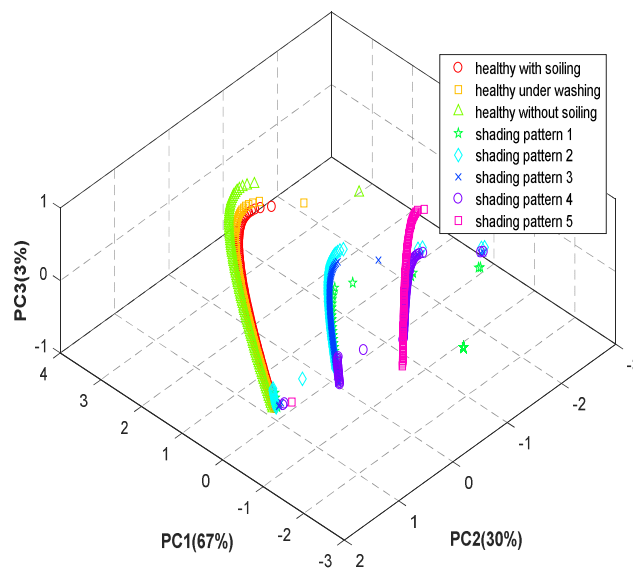Fig. 8: PCA results in the subspace spanned by PC1 and PC2



Fig. 9: PCA results in the subspace spanned by PC1, PC2 and PC3

The previous results are confirmed by the confusion matrix given in table II. The one-leave-out cross validation technique is used to estimate the error rates of class separability into the PCA space. To assign an observation to a class C, we need to measure the Euclidean distance between this observation and the gravity center of C. We first calculate the coordinates of the gravity centers relative to the three obtained classes. Each observation in the dataset is then assigned to a class C among the three classes if this data is the closest to its gravity center. This matrix shows that 98.02% of the data corresponding to the healthy condition are correctly classified. Only 1.93% of these data are misclassified into the faulty class C3. This result

confirms that the PCA highly performs in term of fault detection. For the first faulty class C2, 91.1% of the data are perfectly classified; about 5% and 4% of the data are misclassified respectively into C1 and C3. 88% of the second faulty class C3 data, are well classified but 12% are gathered into C2. This analysis confirms that the classification obtained by the PCA algorithm is efficient.

TABLE II.  CONFUSION MATRIX FOR FAULT CLASSIFICATION ANALYSIS

| | | A posteriori | | |
| --- | --- | --- | --- | --- |
| | | Fault Class C1 | Fault Class C2 | Fault Class C3 |
| A priori | Fault Class C1 | 98.02 % | 0 % | 1.98 % |
| | Fault Class C2 | 4.99 % | 91.1 % | 3.91 % |
| | Fault Class C3 | 0 % | 11.99 % | 88.01% |

## IV. CONCLUSION

A very simple method for the detection of PV cells shading has been presented in this paper. The approach is data-driven and is based on the measured current and voltage. An experimental setup has allowed the collection of data for 8 different operating conditions. After processing the data, the Principal Component Analysis (PCA) is used to analyze the I-V characteristic. The implicit model obtained from the PCA has been used as a new subspace into which the data is projected. The results show a good classification capability relevant with the I-V characteristics. The data are effectively separated with a maximum error of 12% to three different classes: a healthy class and a two faulty classes. Moreover, the proposed method is insensitive to the sudden variations of solar irradiation. In future works, the intra-classes classification will be studied. The performances of the fault diagnosis will also be evaluated in terms of detection and false alarms probability rates.

## REFERENCES

[1]  International Renewable Energy Agency (IRENA), RENEWABLE CAPACITY STATISTICS 2017.

[2]  John H. Wohlgemuth; Sarah Kurtz, "Using accelerated testing to detect module reliability", 37th IEEE Photovoltaic Specialists Conference (PVSC), USA, 2011, pp. 3601-3605.

[3]  F.Bayrak, G. Ertürk, H.F.Oztop, "Effects of shading on energy and exergy efficiencies for photovoltaic panels", Journal of Cleaner Production, 2017, pp. 58-69.

[4]  C.Deline and B.Marion, "A Performance and Economic Analysis of Distributed Power Electronics in Photovoltaic Systems", National Renewable Energy Laboratory (NREL), prepared under task No.PVD9.1410

[5]  J.A.Tsanakas, L.Ha,C.Buerhop, "Faults and infrared thermographic diagnosis in operating C-Si photovoltaic modules: A review of research and future challenges", 2016, Journal of Renewable ans Sustainable Energy Reviews, pp. 695-709

[6]  Y. El Basri, M. Bressan, L. Seguier, H. Alawadhi, C. Alonso "A proposed graphical electrical signatures supervision method to study PV module failures", Journal of Solar energy, 2015, pp 247-256.

[7]  Adria E. Brooksl, Daniel Cormodel, Alexander D. Cronin I, Elsa KamLum, "PV System Power Loss and Module Damage due to Partial Shade and Bypass Diode Failure Depend on Cell Behavior in Reverse Bias", IEEE 2nd Photovoltaic Specialist Conference (PVSC), 2015.

[8]  M.Hassan Ali, A. Hamid Rabhi, A. El hajjaji and Giuseppe M. Tina, "Real Time Fault Detection in Photovoltaic Systems", 8th International Conference on Sustainability in Energy and Buildings, SEB-16, 11-13, September 2016, Turin, ITALY, pp. 914-923.

[9]  E. Garoudja, F. Harrou, Y. Sun, K. Kara, A. Chouder, S. Silvestre, "Statistical fault detection in photovoltaic systems", 6th International Conferenceon Systemsand Control (ICSC), 2017 ,pp.75-80.

[10]  T. Takashima et al., "Experimental studies of fault location in PV module strings," Sol. Energy Mater. Sol. Cells, 2009, pp. 1079–1082.

[11]  T. Takashima et al., "Fault Detection by signal response in PV module Strings", 33rd IEEE Photovoltaic Specialists Conference, 11-16 May 2008, pp.1-5.

[12]  M. K. Alam et al., "PV ground-fault detection using spread spectrum time domain reflectometry (SSTDR)", in Proc. IEEE Energy Convers. Cong.Expo. Sep. 15–19, 2013, pp. 1015–1020.

[13]  M. K. Alam et al., "PV Arc-fault Detection using Spread Spectrum Time Domain Reflectometry (SSTDR) ",  IEEE Energy Conversion Congress and Exposition (ECCE), 14-18 Sept. 2014, pp.3294 – 330.

[14]  M. Bressan , Y. El Basri, A.G. Galeano, C. Alonso "A shadow fault detection method based on the standard error analysis of I-V curves", Journal of Renewable energy, 2016, pp 1181-1190.

[15]  M. Dhimish, V. Holmes, B. Mehradadi, M.Dales, "Diagnostic method for photovoltaic systems based on six layer detection algorithm", Journal of Electric Power Systems Research, 2017, pp 26-39.

[16]  S. Spataru, D. Sera, T. Kerkes, R. Teodorescu, "Diagnostic method for photovoltaic systems based on light I-V measurements", Journal of Solar Energy, 2015, pp 29-44.

[17]  J. Harmouche, C.Delpha, D.Diallo, "Fault diagnosis and detection using Principal Component Analysis and Kullback-Leibler Divergence", 38th Annual Conference on IEEE Industrial Electronics Society (IECON), 25-28 Oct.2012, Montreal, QC, Canada, pp 3907 -3912.

[18]  Gu, Y, Liu, Y & Zhang, Y, "A Selective Kernel PCA Algorithm for Anomaly Detection in Hyperspectral Imagery", IEEE Int. ConI on Acoustics Speed and Signal Processing Proc., vol. 2, pp. 11-725-11-728, May, 2006.

[19]  M. F. Harkat, G. Mourot, J. Ragot, "An improved PCA scheme for sensor FDI: application to an air quality monitoring network", Journal of Process Control, vol. 16, pp. 625-634, 2006.

[20]  M. Haeffelin, L. Barthès, O. Bock, C. Boitel, S. Bony, D. Bouniol, H. Chepfer, M. Chiriaco, J. Cuesta, J. Delanoë, P. Drobinski, J.-L. Dufresne, C. Flamant, M. Grall, A. Hodzic, F. Hourdin, F. Lapouge, Y. Lemaître, A. Mathieu, Y. Morille, C. Naud, V. Noël, W. O'Hirok, J. Pelon, C. Pietras, A. Protat, B. Romand, G. Scialom, and R. Vautard, "SIRTA, a ground-based atmospheric observatory for cloud and aerosol research", Ann. Geophys., vol. 23, no. 2, pp. 253–275, Feb. 2005.

[21]  Jollife, 1.T "Principal Component Analysis, Second ed., Aberdeen U. K", Springer, 2002.